MATHEMATICS 201-510-LW
Business Statistics
Martin Huard
Fall 2008

XXIII – Chi-Square Test of Independence
SOLUTIONS

1. A random sample of 1000 was asked whether they voted in the last general election. Here are the results, broken down by age group.

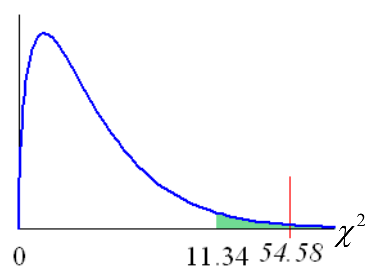
	18-24 years	25-39 years	40-59 years	60 years or older	*Total*
Voted	46 73.7	159 183.2	178 160.4	159 124.7	542
Did not vote	90 62.3	179 154.8	118 135.6	71 105.3	458
Total	136	338	296	230	1000

At the 1% level of significance, test the claim that voting status and are age independent. Try with both approaches, the classical and the *p*-value.

Classical approach

Step 1 Assumptions: The classes are all inclusive and mutually exclusive.

Step 2 H_o: Voting status is independent of age group.
H_A: Voting status is dependent of age group.



Step 3 a) Test statistic: χ^2 with $df = (1)(3) = 3$

b) Right-tailed test with $\alpha = 0.01$

c) $\chi^2_{(df,\alpha)} = \chi^2_{(3,0.01)} = 11.34$

Step 4

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(46-73.7)^2}{73.7} + \frac{(159-183.2)^2}{183.2} + \cdots + \frac{(71-105.3)^2}{105.3}$$

$$= 54.58$$

Step 5 a) χ^2 is in the critical region

b) Reject H_o.

∴ There is sufficient evidence at the 1% level of significance to conclude that voting status is dependent of age group.

***p*-value approach**

Step 1 Assumptions: The classes are all inclusive and mutually exclusive

Step 2 H_o: Voting status is independent of age group.

H_A: Voting status is dependent of age group.

Step 3 a) Test statistic: χ^2 with $df = (1)(3) = 3$

b) Right-tailed test with $\alpha = 0.01$

Step 4 a) $\chi^2 = \sum \frac{(O-E)^2}{E}$

$$= \frac{(25-19.52)^2}{19.52} + \frac{(5-8)^2}{8} + \cdots + \frac{(12-9.52)^2}{9.52}$$

$$= 5.94$$

b) $p\text{-value} < 0.005$

Step 5 a) $p\text{-value} < 0.005 < \alpha = 0.01$

b) Reject H_o.

$\therefore$ There is sufficient evidence at the 1% level of significance to conclude that the two attributes are dependent.

2. A company operates four machines on two separate shifts daily. The following table gives the number of machine breakdowns recorded in the past 6 months.

	Shift 1	Shift 1	*Total*
Machine 1	47 39.3	96 103.7	143
Machine 2	56 47.5	117 125.5	173
Machine 3	27 29.7	81 78.3	108
Machine 4	17 30.5	94 80.5	111
Total	147	388	535

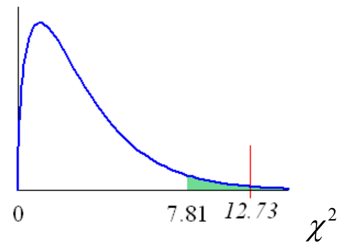
At the 5% level of significance, test whether the machine causing the breakdown is independent of the shift where the breakdown occurred. Try with both approaches, the classical and the *p*-value.

Classical approach

Step 1 Assumptions: The classes are all inclusive and mutually exclusive.

Step 2 H_o: The machine causing the breakdown is independent of the shift where the breakdown occurred.

H_A: The machine causing the breakdown is not independent of the shift where the breakdown occurred.



Step 3 a) Test statistic: χ^2 with $df = (1)(3) = 3$

b) Right-tailed test with α = 0.05

d) $\chi^2_{(df,\alpha)} = \chi^2_{(3,0.05)} = 7.81$

Step 4

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(47-39.3)^2}{39.3} + \frac{(96-103.7)^2}{103.7} + \cdots + \frac{(94-80.5)^2}{80.5}$$

$$= 12.73$$

Step 5 a) χ^2 is in the critical region

b) Reject H_o.

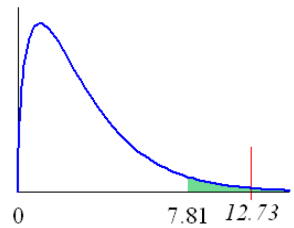
∴ There is sufficient evidence at the 5% level of significance to conclude that the machine causing the breakdown is not independent of the shift where the breakdown occurred.

***p*-value approach**

Step 1 Assumptions: The classes are all inclusive and mutually exclusive

Step 2 H_o: The machine causing the breakdown is independent of the shift where the breakdown occurred.

H_A: The machine causing the breakdown is not independent of the shift where the breakdown occurred.



Step 3 a) Test statistic: χ^2 with $df = (1)(3) = 3$

b) Right-tailed test with α = 0.05

Step 4

a) $\chi^2 = \sum \frac{(O-E)^2}{E}$

$$= \frac{(47-39.3)^2}{39.3} + \frac{(96-103.7)^2}{103.7} + \cdots + \frac{(94-80.5)^2}{80.5}$$

$$= 12.73$$

b) $0.005 < p\text{-value} < 0.01$

Step 5 a) $p\text{-value} < 0.01 < \alpha = 0.05$

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that the machine causing the breakdown is not independent of the shift where the breakdown occurred.

3. A wedding planner took a random sample of 300 men and women were asked "How long should couples date before getting married?" Here are the results.

	Less Than 1 Year	1 Year	1-2 Years	2-3 Years	Longer Than 3 Years	*Total*
Men	31	45	48	16	10	150
	30	44	49.5	16	10.5	
Women	29	43	51	16	11	150
	30	44	49.5	16	10.5	
	60	88	99	32	21	300

At the 5% level of significance, test the claim that a person's response to this question is independent of the person's gender. Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: The classes are all inclusive and mutually exclusive.

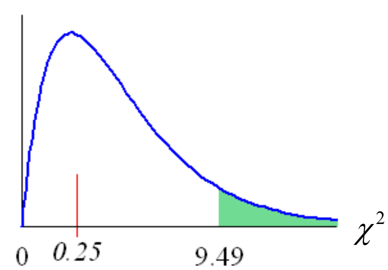
Step 2 H_o: Responses are independent of gender.

H_A: Responses are dependent of gender.

Step 3 a) Test statistic: χ^2 with $df = (1)(4) = 4$

b) Right-tailed test with $\alpha = 0.05$

c) $\chi^2_{(df,\alpha)} = \chi^2_{(4,0.05)} = 9.49$



Step 4

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(31-30)^2}{30} + \frac{(45-44)^2}{44} + \cdots + \frac{(11-10.5)^2}{10.5}$$

$$= 0.25$$

Step 5 a) χ^2 is not in the critical region

b) Fail to reject H_o.

$\therefore$ There is not sufficient evidence at the 5% level of significance to conclude that a person's response to this question is dependent of the person's gender.

***p*-value approach**

Step 1 Assumptions: The classes are all inclusive and mutually exclusive

Step 2 H_o: Responses are independent of gender.

H_A: Responses are dependent of gender.

Step 3 a) Test statistic: χ^2 with $df = (1)(4) = 4$

b) Right-tailed test with α = 0.05

Step 4

a) $\chi^2 = \sum \frac{(O-E)^2}{E}$

$$= \frac{(31-30)^2}{30} + \frac{(45-44)^2}{44} + \cdots + \frac{(11-10.5)^2}{10.5}$$

$= 0.25$

b) $0.990 < p\text{-value} < 0.995$

Step 5 a) $p\text{-value} > 0.990 > \alpha = 0.05$

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 5% level of significance to conclude that a person's response to this question is dependent of the person's gender.

4. An employer wants to determine if the is a relationship between an employee's performance in the company's training program and his or her success on the job (as determined by the employer). A random sample of employees was, and the following results obtained.

		Performance in Training Program		
		Below average	Average	Above Average
Success in job	Excellent	18	36	20
	Very Good	25	44	15
	Good	85	73	82
	Poor	17	12	8

Using a 2.5% significance level, can we conclude that the performance in training program and success in job are dependent? Try with both approaches, the classical and the p-value.

		Performance in Training Program			*Total*
		Below average	Below average	Below average	
Success in job	Excellent	6 26.0	36 34.9	48 29.0	90
	Very Good	25 30.68	44 41.15	37 34.17	106
	Good	30 39.07	61 52.41	44 43.52	135
	Poor	62 27.2	24 36.5	8 30.3	94
	Total	123	165	137	435

Using a 2.5% significance level, can we conclude that the grades are independent of the professor? Try with both approaches, the classical and the p-value.

Classical approach

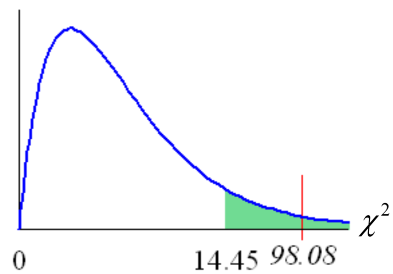
Step 1 Assumptions: The classes are all inclusive and mutually exclusive.

Step 2 H_o: Performance in job training is independent of success on job.
H_A: Performance in job training is dependent of success on job.

Step 3 a) Test statistic: χ^2 with $df = (3)(2) = 6$

b) Right-tailed test with $\alpha = 0.025$

c) $\chi^2_{(df,\alpha)} = \chi^2_{(6,0.025)} = 14.45$



Step 4

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$
$$= \frac{(6-26.0)^2}{26.0} + \frac{(36-34.9)^2}{34.9} + \cdots + \frac{(8-30.3)^2}{30.3}$$
$$= 98.08$$

Step 5 a) χ^2 is in the critical region

b) Reject H_o.

∴ There is sufficient evidence at the 2.5% level of significance to conclude that performance in job training is dependent of success on job.

***p*-value approach**

Step 1 Assumptions: The classes are all inclusive and mutually exclusive

Step 2 H_o: Performance in job training is independent of success on job.
H_A: Performance in job training is dependent of success on job.

Step 3 a) Test statistic: χ^2 with $df = (3)(2) = 4$

b) Right-tailed test with $\alpha = 0.025$

Step 4

a) $$\chi^2 = \sum \frac{(O-E)^2}{E}$$
$$= \frac{(6-26.0)^2}{26.0} + \frac{(36-34.9)^2}{34.9} + \cdots + \frac{(8-30.3)^2}{30.3}$$
$$= 98.08$$

b) p-value < 0.005

Step 5 a) p-value $< 0.005 < \alpha = 0.025$

b) Reject H_o.

∴ There is sufficient evidence at the 2.5% level of significance to conclude that performance in job training is dependent of success on job.

5. As part of a marketing research, a department store compared the income distribution of its shoppers in three locations. Random sample of shoppers were taken at the three locations, and their income noted. Here are the results.

	Income				
	Under $15 000	$15 000 to $29 999	$30 000 to $44 999	$45 000 or more	*Total*
Location 1	91 90.1	82 76.8	74 67.4	40 52.7	287
Location 2	109 114.9	91 98.0	79 85.9	87 67.3	366
Location 3	133 128.1	111 109.2	96 95.8	68 75.0	408
Total	333	284	249	195	1061

Using a 5% level of significance, test the claim that the income of shoppers is independent of location. Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: The classes are all inclusive and mutually exclusive.

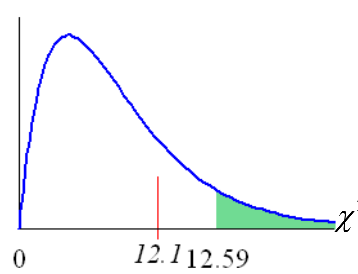
Step 2 H_o: Income is independent of location.

H_A: Income is independent of location.

Step 3 a) Test statistic: χ^2 with $df = (2)(3) = 6$

b) Right-tailed test with $\alpha = 0.05$

c) $\chi^2_{(df,\alpha)} = \chi^2_{(6,0.05)} = 12.59$



Step 4

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(91-90.1)^2}{90.1} + \frac{(82-76.8)^2}{76.8} + \cdots + \frac{(68-75.0)^2}{75.0}$$

$$= 12.10$$

Step 5 a) χ^2 is not in the critical region

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 5% level of significance to conclude that the income of shoppers is dependent of location.

***p*-value approach**

Step 1 Assumptions: The classes are all inclusive and mutually exclusive

Step 2 H_o: Income is independent of location.

H_A: Income is independent of location.

Step 3 a) Test statistic: χ^2 with $df = (2)(3) = 6$

b) Right-tailed test with $\alpha = 0.05$

Step 4 a) $\chi^2 = \sum \frac{(O-E)^2}{E}$

$$= \frac{(91-90.1)^2}{90.1} + \frac{(82-76.8)^2}{76.8} + \cdots + \frac{(68-75.0)^2}{75.0}$$

$$= 12.10$$

b) $0.01 < p\text{-value} < 0.05$

Step 5 a) p-value < $\alpha = 0.05$

b) Reject H_o.

∴ There is not sufficient evidence at the 5% level of significance to conclude that the number of years that a marriage will last is dependent of the highest educational level attained by at least one of the partners.

6. The manager of an assembly process wants to determine whether the number of defective articles manufactured depends on the day of the week the articles are produced. She collected the following information.

Day of the Week	Monday	Tuesday	Wednesday	Thursday	Friday	*Total*
Nondefective	85 91	90 91	95 91	95 91	90 91	455
Defective	15 9	10 9	5 9	5 9	10 9	45
	100	100	100	100	100	500

Using a 10% level of significance, test the claim that the number of defective items is independent of the day of the week. Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: The classes are all inclusive and mutually exclusive.

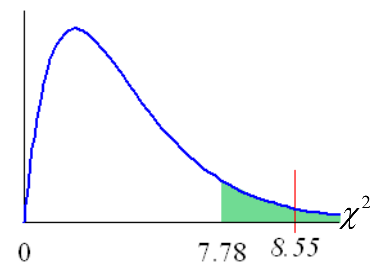
Step 2 Ho: The number of defective items is independent of the day of the week.

H_A: The number of defective items is dependent of the day of the week.

Step 3 a) Test statistic: χ^2 with $df = (1)(4) = 4$

b) Right-tailed test with $\alpha = 0.10$

c) $\chi^2_{(df,\alpha)} = \chi^2_{(4,0.10)} = 7.78$



Step 4

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(85-91)^2}{91} + \frac{(90-91)^2}{91} + \cdots + \frac{(10-9)^2}{9}$$

$$= 8.55$$

Step 5 a) χ^2 is in the critical region

b) Reject H_o.

$\therefore$ There is sufficient evidence at the 5% level of significance to conclude that the number of defective items is dependent of the day of the week.

***p*-value approach**

Step 1 Assumptions: The classes are all inclusive and mutually exclusive

Step 2 Ho: The number of defective items is independent of the day of the week.

H_A: The number of defective items is dependent of the day of the week.

Step 3 a) Test statistic: χ^2 with $df = (1)(4) = 4$

b) Right-tailed test with $\alpha = 0.10$

Step 4

a) $\chi^2 = \sum \frac{(O-E)^2}{E}$

$$= \frac{(85-91)^2}{91} + \frac{(90-91)^2}{91} + \cdots + \frac{(10-9)^2}{9}$$

$$= 8.55$$

b) $0.05 < p\text{-value} < 0.10$

Step 5 a) p-value $> \alpha = 0.05$

b) Reject H_o.

$\therefore$ There is sufficient evidence at the 5% level of significance to conclude that the number of defective items is dependent of the day of the week.